

PENGEMBANGAN INSTRUMEN EVALUASI PEMBELAJARAN PENDIDIKAN AGAMA BUDDHA JENJANG SMA BERBASIS KONTEKSTUAL DAN HOTS PADA KURIKULUM MERDEKA DENGAN PENDEKATAN METODE PILIHAN GANDA KOMPLEKS

Leo Nando Sastro ^{a*)}, Sobani ^{a)}, Shintia Greshia ^{a)}, Yani Relita Sulandari ^{a)}, Juwita Ratna Dewi ^{a)},
Ryan Gautama ^{a)}, Kemanya Karbono ^{a)}, Puji Sulani ^{a)}

^{a)} Sekolah Tinggi Agama Buddha Negeri Sriwijaya, Tangerang Banten, Indonesia

^{*)}e-mail korespondensi: leonandosastro@gmail.com

Article history: received 06 Juni 2026; revised 20 Juni 2026; accepted 17 Juli 2026

DOI : <https://doi.org/10.33751/jmp.v14i02.147>

Abstrak. Penelitian ini bertujuan untuk mengembangkan dan memvalidasi instrumen evaluasi berbasis Keterampilan Berpikir Tingkat Tinggi (HOTS) kontekstual menggunakan item pilihan ganda kompleks untuk mengukur sikap spiritual siswa kelas XI dalam Pendidikan Buddha di bawah Kurikulum Merdeka. Penelitian ini menggunakan pendekatan Penelitian dan Pengembangan (R&D) dengan model ADDIE yang melibatkan 45 siswa dan validator ahli. Kualitas instrumen dievaluasi menggunakan model Rasch, termasuk reliabilitas, validitas konstruk, dimensionalitas, kesesuaian item, dan Fungsi Item Diferensial (DIF). Hasil penelitian menunjukkan sifat psikometrik yang memuaskan, dengan reliabilitas individu 0,77, reliabilitas item 0,98, dan reliabilitas keseluruhan 0,89. Validitas konstruk dikonfirmasi melalui Varians Mentah yang Dijelaskan sebesar 77,2%, sementara hanya satu item yang menunjukkan bias gender. Instrumen yang dikembangkan secara efektif mengukur kemampuan berpikir tingkat tinggi pada tingkat analisis dan evaluasi dalam konteks pembelajaran autentik dan sesuai untuk menilai sikap spiritual siswa dalam Pendidikan Buddha.

Kata Kunci: Keterampilan Berpikir Tingkat Tinggi (HOTS); Model Rasch; Penilaian Sikap Spiritual

DEVELOPMENT OF A CONTEXTUAL AND HOTS-BASED EVALUATION INSTRUMENT FOR BUDDHIST RELIGIOUS EDUCATION LEARNING AT HIGH SCHOOL LEVEL IN THE INDEPENDENT CURRICULUM WITH A COMPLEX MULTIPLE CHOICE METHOD APPROACH

Abstract. This study aimed to develop and validate a contextual Higher Order Thinking Skills (HOTS)-based evaluation instrument using complex multiple-choice items to measure the spiritual attitudes of Grade XI students in Buddhist Education under the Merdeka Curriculum. The research employed a Research and Development (R&D) approach with the ADDIE model involving 45 students and expert validators. Instrument quality was evaluated using the Rasch model, including reliability, construct validity, dimensionality, item fit, and Differential Item Functioning (DIF). The results demonstrated satisfactory psychometric properties, with person reliability of 0.77, item reliability of 0.98, and overall reliability of 0.89. Construct validity was confirmed through a Raw Variance Explained of 77.2%, while only one item exhibited gender bias. The developed instrument effectively measures higher-order thinking at the analysis and evaluation levels within authentic learning contexts and is suitable for assessing students' spiritual attitudes in Buddhist Education.

Keywords: Higher Order Thinking Skills (HOTS); Rasch Model; Spiritual Attitude Assessment

I. PENDAHULUAN

Pendidikan merupakan proses sistematis yang bertujuan untuk mengembangkan potensi peserta didik secara optimal, baik dalam ranah kognitif, afektif, maupun psikomotor. Dalam proses tersebut, evaluasi pembelajaran menjadi komponen penting yang tidak dapat dipisahkan, karena berfungsi sebagai alat untuk mengukur ketercapaian tujuan pembelajaran serta sebagai dasar dalam pengambilan keputusan pada proses pembelajaran. Arikunto (2013) menyatakan bahwa evaluasi merupakan kegiatan untuk mengetahui sejauh mana tujuan pembelajaran telah tercapai, sehingga kualitas evaluasi sangat ditentukan oleh kualitas instrumen yang digunakan. Instrumen evaluasi yang baik seharusnya mampu mengukur kemampuan peserta didik secara komprehensif, valid, dan reliabel. Namun, realitas di lapangan menunjukkan bahwa instrumen evaluasi yang digunakan oleh pendidik masih didominasi oleh soal-soal yang mengukur kemampuan berpikir tingkat rendah (Lower Order Thinking Skills/LOTS), seperti mengingat dan memahami. Penelitian Widana (2017) mengungkapkan bahwa sebagian besar instrumen penilaian yang digunakan

pendidik belum mengakomodasi kemampuan berpikir tingkat tinggi (Higher Order Thinking Skills/HOTS), sehingga belum mampu mengukur kemampuan analisis, evaluasi, dan pemecahan masalah peserta didik secara optimal.

Tuntutan pembelajaran abad ke-21 menekankan pentingnya penguasaan kemampuan berpikir tingkat tinggi (Higher Order Thinking Skills/HOTS), yang mencakup kemampuan menganalisis, mengevaluasi, dan mencipta. Anderson dan Krathwohl (2001) dalam revisi Taksonomi Bloom menegaskan bahwa kemampuan berpikir tingkat tinggi merupakan bagian penting dalam pembelajaran bermakna. Hal ini diperkuat oleh penelitian Heong et al. (2011) yang menyatakan bahwa HOTS sangat diperlukan untuk meningkatkan kemampuan berpikir kritis dan pemecahan masalah peserta didik. Selain itu, Mardapi (2012) menegaskan bahwa instrumen penilaian harus dirancang secara sistematis agar mampu menghasilkan data yang valid dan dapat digunakan sebagai dasar pengambilan keputusan. Black dan Wiliam (1998) juga menyatakan bahwa penilaian yang berkualitas dapat meningkatkan proses pembelajaran melalui umpan balik yang konstruktif.

Pengembangan HOTS juga menjadi fokus dalam implementasi kurikulum, termasuk Kurikulum Merdeka. Penelitian oleh Retnawati et al. (2018) menunjukkan bahwa kemampuan HOTS peserta didik di Indonesia masih tergolong rendah, sehingga diperlukan perbaikan dalam sistem pembelajaran dan evaluasi, khususnya dalam penyusunan instrumen penilaian. Oleh karena itu, pendidik dituntut untuk mampu mengembangkan instrumen evaluasi yang tidak hanya mengukur pengetahuan, tetapi juga keterampilan berpikir tingkat tinggi. Permasalahan tersebut menjadi semakin kompleks dalam pembelajaran Pendidikan Agama Buddha dan Budi Pekerti, khususnya pada materi “Sayangi Seni dan Budaya Agama Buddha”. Materi ini tidak hanya menuntut pemahaman konsep, tetapi juga kemampuan peserta didik dalam menganalisis keberagaman budaya, mengevaluasi nilai-nilai toleransi, serta mengaplikasikan ajaran Buddha dalam kehidupan sehari-hari. Suryani (2020) juga menegaskan bahwa pembelajaran berbasis nilai dan budaya memerlukan instrumen evaluasi yang kontekstual agar mampu mengukur sikap dan karakter peserta didik secara menyeluruh. Namun demikian, instrumen yang tersedia belum sepenuhnya mampu mengukur keterampilan tersebut secara mendalam dan terintegrasi. Selain itu, instrumen yang digunakan juga belum sepenuhnya selaras dengan tuntutan Kurikulum Merdeka yang berbasis Capaian Pembelajaran (CP), Tujuan Pembelajaran (TP), dan Alur Tujuan Pembelajaran (ATP). Instrumen evaluasi yang tidak terintegrasi dengan komponen tersebut akan menyebabkan penilaian menjadi kurang optimal. Sudjana (2017) menyatakan bahwa penilaian yang baik harus relevan dengan tujuan pembelajaran serta mampu memberikan gambaran yang komprehensif mengenai hasil belajar peserta didik. Hal ini juga didukung oleh penelitian oleh Mardapi (2012) yang menegaskan bahwa instrumen penilaian harus disusun secara sistematis agar mampu menghasilkan data yang valid dan dapat dipertanggungjawabkan. Penilaian dalam pendidikan tidak hanya berfungsi sebagai alat ukur hasil belajar, tetapi juga sebagai sarana untuk meningkatkan kualitas pembelajaran. Black dan Wiliam (1998) menyatakan bahwa penilaian yang efektif dapat memberikan umpan balik yang konstruktif bagi peserta didik maupun pendidik. Penelitian yang dilakukan oleh Haryanto (2020) menunjukkan bahwa penggunaan instrumen penilaian yang tepat dapat meningkatkan kualitas proses pembelajaran dan hasil belajar peserta didik secara signifikan. Berdasarkan kondisi tersebut, dapat dirumuskan beberapa permasalahan utama, yaitu: (1) instrumen evaluasi masih berfokus pada LOTS dan belum mengukur HOTS secara optimal; (2) belum adanya instrumen yang mampu mengukur aspek kognitif, afektif, dan psikomotor secara terpadu; (3) keterbatasan instrumen dalam mengaitkan konteks pembelajaran dengan kehidupan nyata; serta (4) belum terintegrasinya instrumen dengan komponen Kurikulum Merdeka. Permasalahan ini menunjukkan adanya kesenjangan antara tuntutan kurikulum dengan praktik evaluasi di lapangan.

Berdasarkan urgensi tersebut, pengembangan instrumen evaluasi pembelajaran menjadi suatu kebutuhan yang mendesak, baik secara teoritis maupun praktis. Secara teoritis, pengembangan instrumen ini berkontribusi dalam memperkaya kajian tentang evaluasi pembelajaran berbasis HOTS dan kontekstual. Secara praktis, instrumen ini dapat digunakan oleh pendidik sebagai alat ukur yang lebih akurat dalam menilai kemampuan peserta didik serta sebagai dasar dalam perbaikan proses pembelajaran. Instrumen yang dikembangkan dalam penelitian ini berupa tes pilihan ganda berbasis HOTS dengan model pilihan ganda kompleks, yaitu bentuk soal yang memungkinkan lebih dari satu jawaban benar dalam satu butir soal. Model ini dipilih karena mampu mengukur kemampuan analisis dan evaluasi peserta didik secara lebih mendalam dibandingkan dengan pilihan ganda konvensional. Menurut penelitian oleh Istiyono et al. (2014), instrumen berbasis HOTS yang dirancang dengan baik dapat meningkatkan kemampuan berpikir kritis peserta didik dan memberikan hasil pengukuran yang lebih akurat. Adapun kebaruan dalam penelitian ini terletak pada pengembangan instrumen evaluasi yang: (1) berbasis HOTS pada level kognitif C4 (analisis) dan C5 (evaluasi); (2) menggunakan model pilihan ganda kompleks sebagai bentuk inovasi dalam penilaian; (3) mengintegrasikan konteks budaya lokal dan kehidupan nyata dalam soal; serta (4) disusun secara sistematis berdasarkan CP, TP, dan ATP dalam Kurikulum Merdeka. Dengan demikian, instrumen yang dikembangkan diharapkan mampu menghasilkan penilaian yang lebih komprehensif, objektif, dan relevan dengan kebutuhan pembelajaran abad ke-21.

II. METODE PENELITIAN

Jenis penelitian ini menggunakan metode *Research and Development* (R&D) dimana penelitian ini bertujuan untuk menghasilkan produk berupa instrumen evaluasi pembelajaran berbasis Higher Order Thinking Skills (HOTS) dalam bentuk soal pilihan ganda kompleks untuk mengukur sikap spiritual peserta didik pada mata pelajaran Pendidikan Agama Buddha dan Budi Pekerti. Menurut Sugiyono (2022), penelitian R&D merupakan metode penelitian yang digunakan untuk menghasilkan produk tertentu serta menguji keefektifan produk tersebut. Sejalan dengan itu, penelitian oleh Haryati (2021) menyatakan bahwa R&D sangat relevan digunakan dalam pengembangan instrumen pembelajaran karena mampu menghasilkan produk yang valid, praktis,

dan efektif melalui tahapan sistematis. Penelitian ini dilakukan di Jakarta pada bulan Februari sampai dengan dengan bulan Juli 2026. Subjek penelitian dan perkembangan instrumen menggunakan teknik *purpose sampling* dengan sample siswa kelas XI, berjumlah 45 orang. Penelitian ini melibatkan guru sebagai penulis dan pembuat soal serta pakar (Ahli Matrei Pendidikan Agama Buddha) penelitian yang bertugas untuk mengevaluasi dan pengembangan soal.

Model pengembangan yang digunakan dalam penelitian ini adalah model *ADDIE* (Analysis, Design, Development, Implementation, Evaluation). Model ini dipilih karena bersifat sistematis, fleksibel, dan banyak digunakan dalam pengembangan perangkat pembelajaran. Menurut Branch (2020), model *ADDIE* terdiri dari lima tahapan utama, yaitu; tahap pertama, Tahap ini dilakukan untuk mengidentifikasi kebutuhan pembelajaran, karakteristik peserta didik, serta permasalahan dalam evaluasi pembelajaran. Analisis juga mencakup kajian terhadap Capaian Pembelajaran (CP), Tujuan Pembelajaran (TP), dan Alur Tujuan Pembelajaran (ATP) dalam Kurikulum Merdeka. Tahap kedua adalah *design* yaitu, tahap ini dilakukan perancangan instrumen evaluasi, meliputi penyusunan kisi-kisi soal, indikator HOTS (C4 dan C5), serta pengembangan bentuk soal pilihan ganda kompleks berbasis situasi kontekstual. Tahap ketiga adalah pengembangan yang merupakan proses pembuatan instrumen sesuai dengan desain yang telah disusun. Instrumen kemudian divalidasi oleh para ahli (expert judgment) untuk mengetahui tingkat validitas isi dan konstruk. Tahap keempat adalah implementasi dimana Instrumen yang telah dinyatakan valid oleh pakar kemudian diuji cobakan kepada peserta didik untuk mengetahui kepraktisan dan keefektifannya dalam mengukur sikap spiritual. Dan tahap yang terakhir adalah Tahap ini dilakukan untuk mengevaluasi kualitas instrumen berdasarkan hasil uji coba, meliputi analisis validitas, reliabilitas, tingkat kesukaran, dan daya pembeda soal. Penelitian oleh Cahyadi (2020) menyatakan bahwa model *ADDIE* efektif digunakan dalam pengembangan instrumen pembelajaran berbasis HOTS karena mampu menghasilkan produk yang sistematis dan teruji.

III. HASIL DAN PEMBAHASAN

Hasil dan Pengolahan Data dengan menggunakan Simulasi Analisis Data Politomi (Skala Mindfull Speech) dengan pemodelan *RASCH* atas pelaksanaan penelitian yang dilakukan di Jakarta pada bulan Februari sampai dengan dengan bulan Juli 2026 dengan subjek penelitian dan perkembangan instrumen sudah dijelaskan pada Bab III tersebut di atas dengan menggunakan teknik *purpose sampling* dengan sample siswa kelas XI, berjumlah 45 orang dan juga melibatkan guru sebagai penulis dan pembuat soal serta pakar (Ahli Materi Pendidikan Agama Buddha) penelitian yang bertugas untuk mengevaluasi dan pengembangan soal termasuk model pengembangan yang digunakan dalam penelitian ini adalah model *ADDIE* (Analysis, Design, Development, Implementation, Evaluation) yang bersifat sistematis, fleksibel, dan banyak digunakan dalam pengembangan perangkat pembelajaran. Model pengembangan yang digunakan dalam penelitian ini adalah model *ADDIE* (Analysis, Design, Development, Implementation, Evaluation) dan pengolahan data Simulasi Analisis Data Politomi (Skala Mindfull Speech) Dengan Pemodelan *Rasch*.

TABEL 1 SUMMARY STATISTIC – 3.1

Menjelaskan terkait informasi secara keseluruhan tentang kualitas responden, kualitas instrumen dan interaksi antara person dan item, yang dapat dijelaskan dan uraikan dalam table di bawah ini, sebagai berikut:

Tabel 1 Summary Statistic – 3.1

TABLE 3.1 C:\Users\DELL\Desktop\Data pilgan (wit ZOU274NS.TXT, Jun 23 14:44 2026
 INPUT: 45 Person 30 Item REPORTED: 45 Person 30 Item 6 CATS WINSTEPS 3.73

SUMMARY OF 45 MEASURED Person									
	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	
MEAN	168.7	30.0	2.28	.45	.74	-.4	.66	-.6	
S.D.	8.1	.0	.99	.11	.50	1.3	.84	1.2	
MAX.	168.0	30.0	3.66	.60	2.14	3.7	5.15	3.7	
MIN.	139.0	30.0	-.05	.26	.09	-2.0	.05	-2.0	
REAL RMSE	.47	TRUE SD	.67	SEPARATION	1.84	Person	RELIABILITY	.77	
MODEL RMSE	.46	TRUE SD	.68	SEPARATION	1.90	Person	RELIABILITY	.78	
S.E. OF Person MEAN = .15									
Person RAW SCORE-TO-MEASURE CORRELATION = .97									
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .89									
SUMMARY OF 30 MEASURED Item									
	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	
MEAN	241.1	45.0	.00	.36	.89	-.3	.66	-.8	
S.D.	33.9	.0	2.56	.09	.47	1.5	.64	1.2	
MAX.	385.0	45.0	3.14	.49	1.93	2.3	3.40	2.5	
MIN.	205.0	45.0	-4.78	.20	.08	-5.1	.05	-2.8	
REAL RMSE	.39	TRUE SD	2.53	SEPARATION	6.43	Item	RELIABILITY	.98	
MODEL RMSE	.37	TRUE SD	2.53	SEPARATION	6.78	Item	RELIABILITY	.98	
S.E. OF Item MEAN = .48									

Dalam tabel *Summary Statistic* di atas, menunjukkan banyaknya data person dan item yang diolah adalah 45 Person dengan 30 Item. Data analisa yang dapat diberikan, adalah *Person Measure*, menunjukkan nilai rata-rata logit responden dalam *mindful speech scale*. Nilai Rata-rata > 0.0 menunjukkan kecenderungan responden yang lebih banyak menjawab benar dari 30 soal pada pertanyaan-pertanyaan Pilihan Ganda Kompleks. Kesesuaian data (INFIT dan OUTFIT Statistic) terhadap model *RASCH*, apakah pola jawaban atau item sesuai dengan yang diprediksi model *Rasch*, berikut penjelasannya, Untuk *MNSQ* pada tabel person nilainya 0.74 dan 0.66 (idealnya adalah 0, artinya semakin mendekati 0, maka semakin bagus hasilnya) dan Untuk *ZSTD* nilainya -0.4 dan -0.6 (idealnya adalah 0, artinya semakin mendekati 0, maka kualitasnya semakin baik). Pengelompokan Person/Orang ditunjukkan pada *Separation* dengan nilai 1.84, semakin besar nilai ini maka kualitas instrument dalam keseluruhan responden dan item semakin baik. Bisa digunakan untuk pengelompokan. Dengan formula H, maka person dapat dikelompokkan menjadi dua. Nilai Person *Reliability*, Nilai 0.77 dalam kategori reliabilitas kuat (<0,67) artinya konsistensi jawaban para siswa kuat/bagus. Reliabilitas interaksi antara *Person* dan Item secara keseluruhan (Internal Consistency), menunjukkan nilai 0.89 yang artinya dalam kategori baik/bagus. *Error Standar Item*, semakin kecil maka semakin baik dengan mean item adalah 0.0 dan S.E. of Item Mean adalah 0.48. Artinya rata-rata kesulitan item berada pada 0 logit dengan kesalahan estimasi ± 0.89

Pengelompokan Item, dengan formula

$$H = \frac{((4 \times Separation) + 1)}{3} = \frac{((4 \times 6.43) + 1)}{3} = 9$$

Artinya, item dapat dikelompokkan kedalam 9 kelompok, (bisa dari sangat sangat mudah, sangat mudah, sampai sangat sangat sukar) > 9 kelompok. Ini berlaku juga untuk *Separation Person*. Nilai reliability sebesar 0,98 kategori Istimewa, artinya kualitas Item dalam Instrument bagus.

TABEL 2 ITEM MEASURE – 13.1

Nilai logit setiap item yang diurutkan dari tertinggi ke terendah, sebagai berikut, Item soal paling sulit adalah soal nomor 5, paling sulit dijawab benar oleh responden dan Item soal paling mudah adalah soal nomor 14, dengan nilai logit -4.78. Label item soal yang diurutkan dari soal yang dikerjakan paling sulit sampai dengan soal yang dikerjakan paling mudah (*ditunjukkan pada kolom paling kanan dengan judul kolom Item*). Standar Deviasi logit dapat digunakan untuk pengelompokan item berdasarkan nilai acuan norma dengan ditunjukkan oleh *MEAN S.D.* 0.00 – 0.48.

Tabel 2 Item Measure – 13.1

TABLE 13.1 C:\Users\DELL\Desktop\Data **gila** (g) ZOU274W5.TXT Jun 23 14:44 2826
 INPUT: 45 Person 30 Item REPORTED: 45 Person 30 Item 6 CATS WINSTEPS 3.73

Person: REAL SEP.: 1.84 REL.: .77 ... Item: REAL SEP.: 6.43 REL.: .98

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MODEL MEASURE	INFIT S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	EXACT MATCH EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
5	285	45	3.14	.26	1.09	.4	.99	.1	.40	.63	55.6	73.3	S5
8	286	45	2.92	.28	.92	-.2	.61	-1.0	.70	.61	68.9	74.6	S8
17	215	45	2.30	.33	1.02	.2	.38	-1.4	.66	.52	80.0	79.0	S17
4	217	45	2.08	.35	.66	-1.0	.73	-.3	.48	.48	84.4	83.1	S4
3	218	45	1.95	.36	.60	-1.1	.45	-1.0	.57	.46	86.7	83.8	S3
16	218	45	1.95	.36	1.11	.4	.60	-.6	.44	.46	84.4	83.8	S16
19	218	45	1.95	.36	.92	-.1	.75	-.3	.42	.46	84.4	83.8	S19
1	219	45	1.81	.38	.40	-1.9	.13	-2.2	.69	.43	95.6	85.3	S1
2	219	45	1.81	.38	.49	-1.5	.16	-2.1	.66	.43	91.1	85.3	S2
7	220	45	1.66	.40	1.13	.4	.44	-.9	.43	.41	91.1	86.8	S7
18	220	45	1.66	.40	1.00	.1	.52	-.7	.43	.41	88.9	86.8	S18
20	221	45	1.50	.42	.88	-.1	.23	-1.6	.48	.38	93.3	87.4	S20
30	222	45	1.31	.44	1.93	1.6	3.40	2.5	-.01	.35	97.8	90.7	S30
21	223	45	1.11	.47	.37	-1.4	.17	-1.8	.38	.32	95.6	91.5	S21
24	223	45	1.11	.47	.72	-.4	.22	-1.6	.32	.32	97.8	91.5	S24
27	223	45	1.11	.47	.68	-.5	.19	-1.7	.34	.32	97.8	91.5	S27
28	223	45	1.11	.47	.68	-.5	.19	-1.7	.34	.32	97.8	91.5	S28
29	223	45	1.11	.47	.39	-1.3	.39	-1.0	.27	.32	95.6	91.5	S29
22	224	45	.88	.48	.19	-2.0	.07	-2.4	.34	.29	97.8	92.1	S22
25	224	45	.88	.48	.19	-2.0	.07	-2.4	.34	.29	97.8	92.1	S25
26	225	45	.64	.49	.06	-2.7	.05	-2.8	.00	.26	100.0	92.2	S26
6	263	45	-2.30	.28	.42	-5.1	1.06	.3	.69	.51	77.0	40.8	S6
15	283	45	-3.18	.23	.93	-.3	.89	-.3	.64	.62	62.2	58.1	S15
11	289	45	-3.51	.24	1.69	2.3	1.62	1.7	.45	.64	55.6	66.1	S11
12	291	45	-3.63	.25	1.05	.3	.87	-.3	.76	.64	73.3	69.3	S12
9	295	45	-3.90	.27	1.03	.2	.95	.0	.59	.64	64.4	74.4	S9
10	298	45	-4.13	.28	1.41	1.4	.74	-.5	.67	.62	80.0	75.2	S10
13	300	45	-4.29	.29	1.27	1.0	.57	-.9	.73	.60	84.4	76.0	S13
23	300	45	-4.29	.29	1.72	2.1	1.30	.7	.45	.60	75.6	76.0	S23
14	305	45	-4.78	.33	1.64	1.8	1.11	.4	.41	.53	73.3	79.6	S14
MEAN	241.1	45.0	.00	.36	.89	-.3	.66	-.8			84.3	81.1	
S.D.	33.9	.0	2.56	.09	.47	1.5	.64	1.2			13.0	11.4	

Nilai logit setiap item yang diurutkan dari tertinggi ke terendah, sebagai berikut, Item soal paling sulit adalah soal nomor 5, paling sulit dijawab benar oleh responden dan Item soal paling mudah adalah soal nomor 14, dengan nilai logit -4.78. Label item soal yang diurutkan dari soal yang dikerjakan paling sulit sampai dengan soal yang dikerjakan paling mudah (*ditunjukkan*

pada kolom paling kanan dengan judul kolom Item). Standar Deviasi logit dapat digunakan untuk pengelompokan item berdasarkan nilai acuan norma dengan ditunjukkan oleh *MEAN S.D.* 0.00 – 0.48.

TABEL 3 TABLE ITEM FIT ORDER – 10.1

Tabel 3 Tabel Item Fit Order – 10.1

TABLE 10.1 C:\Users\DELL\Desktop\Data pilaep (uji ZOU27445.TXT) Jun 23 14:44 2026
INPUT: 45 Person 30 Item REPORTED: 45 Person 30 Item 6 CATS WINGSTEPS 3.73
Person: REAL SEP.: 1.84 REL.: .77 ... Item: REAL SEP.: 6.43 REL.: .98

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFI T	OUTFI T	PT-MEASU RE CORR.	EXACT MATCH	Item
					MNSQ	ZSTD	MNSQ	ZSTD	
30	222	45	1.31	.44	1.93	1.6	3.40	2.5	A-.01
23	300	45	-4.29	.29	1.72	2.1	1.30	.7	B-.45
11	289	45	-3.51	.24	1.69	2.3	1.62	1.7	C-.45
14	305	45	-4.78	.33	1.64	1.8	1.11	.4	D-.41
10	298	45	-4.13	.28	1.41	1.4	.74	-.5	E-.67
13	300	45	-4.29	.29	1.27	1.0	.57	-.9	F-.73
7	220	45	1.66	.40	1.13	.4	.44	-.9	G-.43
16	218	45	1.95	.36	1.11	.4	.60	-.6	H-.44
5	205	45	3.14	.26	1.09	.4	.99	.1	I-.40
6	263	45	-2.30	.20	.42	-5.1	1.06	.3	J-.69
12	291	45	-3.63	.25	1.05	.3	.87	-.3	K-.76
9	295	45	-3.90	.27	1.03	.2	.95	.0	L-.59
17	215	45	2.30	.33	1.02	.2	.38	-1.4	M-.66
18	220	45	1.66	.40	1.00	.1	.52	-.7	N-.43
15	283	45	-3.18	.23	.93	-.3	.89	-.3	O-.64
19	218	45	1.95	.36	.92	-.1	.75	-.3	o-.42
8	208	45	2.92	.28	.92	-.2	.61	-1.0	n-.70
20	221	45	1.50	.42	.88	-.1	.23	-1.6	m-.48
4	217	45	2.08	.35	.66	-1.0	.73	-.3	l-.48
24	223	45	1.11	.47	.72	-.4	.22	-1.6	k-.32
27	223	45	1.11	.47	.68	-.5	.19	-1.7	j-.34
28	223	45	1.11	.47	.68	-.5	.19	-1.7	i-.34
3	218	45	1.95	.36	.60	-1.1	.45	-1.0	h-.57
2	219	45	1.81	.38	.49	-1.5	.16	-2.1	g-.66
1	219	45	1.81	.38	.40	-1.9	.13	-2.2	f-.69
29	223	45	1.11	.47	.39	-1.3	.39	-1.0	e-.27
21	223	45	1.11	.47	.37	-1.4	.17	-1.8	d-.38
22	224	45	.88	.48	.19	-2.0	.07	-2.4	c-.34
25	224	45	.88	.48	.19	-2.0	.07	-2.4	b-.34
26	225	45	.64	.49	.08	-2.7	.05	-2.8	a-.00
MEAN	241.1	45.0	.00	.36	.09	-.3	.66	-.8	
S.D.	33.9	.0	2.56	.09	.47	1.5	.64	1.2	

Tiga kolom (Outfit Mnsq, Outfit Zstd Dan Pt Measure Corr), merupakan kriteria kesesuaian item (Outliers atau Misfit). Kriteria yang digunakan, adalah sebagai berikut, *Outfit MNSQ* diterima $0.5 < MNSQ < 1.5$, *Outfit ZSTD* antara -2.0 sampai +2.0. *PT Measure Corr.* 0.4 sampai 0.85. Ini juga dapat digunakan sebagai kriteria item FIT dengan menjumlahkan *MEAN S.D.* lalu dibandingkan dengan nilai *INFI T MNSQ* setiap item. Item yang *INFI T MNSQ* tersebut lebih besar nilainya (>), berarti dapat menjadi kategori *Misfit Order*.

TABEL 4 TABLE PERSON MEASURE ORDER – 17.1

Tabel 4 Tabel Person Measure Order – 17.1

TABLE 17.1 C:\Users\DELL\Desktop\Data **plisan (ki 20U274MS.TXT)** Jun 23 14:44 2026
 INPUT: 45 Person 30 Item REPORTED: 45 Person 30 Item 6 CATS WINSTEPS 3.73
 Person: REAL SEP.: 1.84 REL.: .77 ... Item: REAL SEP.: 6.43 REL.: .98

Person STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	EXP.	EXACT MATCH OBS% EXP%	Person
28	168	30	3.66	.60	.09	-2.0	.05	-2.0	.97	.93	100.0 92.6	28P
6	167	30	3.50	.59	.12	-2.0	.05	-2.0	.97	.92	100.0 91.0	06P
8	167	30	3.50	.59	.12	-2.0	.05	-2.0	.97	.92	100.0 91.0	08P
13	167	30	3.50	.59	.12	-2.0	.05	-2.0	.97	.92	100.0 91.0	13L
18	167	30	3.50	.59	.12	-2.0	.05	-2.0	.97	.92	100.0 91.0	18P
33	167	30	3.50	.59	.12	-2.0	.05	-2.0	.97	.92	100.0 91.0	33P
39	167	30	3.50	.59	.12	-2.0	.05	-2.0	.97	.92	100.0 91.0	39P
2	166	30	2.97	.55	.34	-1.3	.35	-.8	.95	.91	96.7 90.3	02L
3	166	30	2.97	.55	.32	-1.3	.22	-1.2	.95	.91	96.7 90.3	03L
23	166	30	2.97	.55	.64	-.5	.33	-.9	.93	.91	90.0 90.3	23L
27	166	30	2.97	.55	.31	-1.4	.21	-1.3	.97	.91	96.7 90.3	27P
4	165	30	2.70	.50	.85	-.1	.37	-.9	.90	.90	96.7 89.0	04L
7	165	30	2.70	.50	.72	-.4	.24	-1.2	.91	.90	96.7 89.0	07P
11	165	30	2.70	.50	.39	-1.3	.22	-1.3	.94	.90	93.3 89.0	11P
17	165	30	2.70	.50	.88	.0	.43	-.7	.90	.90	96.7 89.0	17L
29	165	30	2.70	.50	.51	-.9	.63	-.3	.96	.90	93.3 89.0	29P
35	165	30	2.70	.50	.50	-.9	.57	-.4	.94	.90	93.3 89.0	35P
36	165	30	2.70	.50	.33	-1.5	.18	-1.4	.97	.90	93.3 89.0	36P
38	165	30	2.70	.50	.66	-.5	.27	-1.1	.93	.90	86.7 89.0	38P
41	165	30	2.70	.50	.39	-1.3	.22	-1.3	.96	.90	93.3 89.0	41L
9	164	30	2.46	.47	.80	-.3	.33	-1.0	.89	.88	90.0 86.1	09L
12	164	30	2.46	.47	.74	-.4	.35	-1.0	.89	.88	93.3 86.1	12L
24	164	30	2.46	.47	.60	-.8	.55	-.5	.90	.88	86.7 86.1	24P
34	164	30	2.46	.47	.66	-.6	.24	-1.3	.90	.88	93.3 86.1	34L
37	164	30	2.46	.47	.49	-1.1	.51	-.6	.91	.88	93.3 86.1	37P
42	164	30	2.46	.47	.90	.0	.63	-.3	.93	.88	80.0 86.1	42L
19	163	30	2.26	.43	.82	-.2	.33	-1.1	.85	.87	86.7 84.6	19P
25	163	30	2.26	.43	.52	-1.1	.52	-.6	.96	.87	90.0 84.6	25P
26	163	30	2.26	.43	1.98	1.8	5.15	3.4	.85	.87	90.0 84.6	26P
45	163	30	2.26	.43	.55	-1.0	.65	-.3	.93	.87	90.0 84.6	45P
14	162	30	2.08	.41	.83	-.3	.33	-1.2	.86	.85	90.0 81.8	14L
21	162	30	2.08	.41	.61	-.9	.65	-.4	.96	.85	83.3 81.8	21L
22	162	30	2.08	.41	.74	-.5	.70	-.3	.92	.85	76.7 81.8	22P
20	161	30	1.93	.39	1.02	.2	.96	.2	.85	.83	80.0 79.1	20L
31	161	30	1.93	.39	.46	-1.5	.37	-1.1	.94	.83	83.3 79.1	31L
50	160	30	1.78	.37	1.66	1.6	1.23	.6	.81	.81	70.0 79.1	50L
16	151	30	.84	.29	1.29	1.1	1.15	.5	.54	.67	63.3 62.4	16L
10	149	30	.68	.28	1.49	1.7	1.16	.5	.49	.65	66.7 60.7	10L
15	149	30	.68	.28	1.56	2.0	1.12	.4	.26	.65	60.0 60.7	15P
40	149	30	.68	.28	.98	.0	.82	-.4	.62	.65	63.3 60.7	40P
5	147	30	.52	.28	1.22	.9	1.17	.6	.45	.64	50.0 55.8	05L
1	144	30	.30	.27	1.09	.4	.88	-.3	.56	.62	60.0 53.6	01L
44	141	30	.09	.26	1.39	1.5	1.75	2.0	.44	.61	46.7 51.7	44L
32	140	30	.02	.26	1.84	.2	1.00	.1	.56	.61	60.0 51.6	32L
43	139	30	-.05	.26	2.14	3.7	2.61	3.7	.73	.61	23.3 53.8	43L
MEAN	160.7	30.0	2.20	.45	.74	-.4	.66	-.6			84.3 81.1	
S.D.	8.1	.0	.99	.11	.50	1.3	.84	1.2			17.0 12.7	

Nilai Logit setiap person yang diurutkan dari abilitas tertinggi sampai dengan yang terendah, Person dengan abilitas tertinggi adalah person dengan no urut 28, jenis kelamin Perempuan dengan nama Chelsea Natasya dengan logit 3.66 dan Person dengan abilitas terendah adalah person dengan no urut 43, jenis kelamin Laki-laki dengan nama Nickholas Andrew Wijaya dengan logit -0.05.

TABEL 5a TABEL PERSON FIT ORDER – 6.1

TABEL 5B GUTTMAN SCALOGRAMS OF RESPONSES – 22.0

Tabel 5a Tabel Person Fit Order – 6.1

TABLE 6.1 C:\Users\DELL\Desktop\Data pilgao (wit ZOU274WS.TXT, Jun 23 14:44 2026
INPUT: 45 Person 30 Item REPORTED: 45 Person 30 Item 6 CATS WINSTEP5 3.73

Person: REAL SEP.: 1.84 REL.: .77 ... Item: REAL SEP.: 6.43 REL.: .98

Person STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MODEL MEASURE	S.E.	INFI MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	EXP.	EXACT MATCH OBS% EXP%	Person
26	163	30	2.26	.43	1.98	1.6	5.15	3.4	.85	.87	90.0 84.6	26P
43	139	30	-.05	.26	2.14	3.7	2.61	3.7	.73	.61	23.3 53.8	43L
44	141	30	.09	.26	1.39	1.5	1.75	2.0	.44	.61	46.7 51.7	44L
30	160	30	1.78	.37	1.66	1.6	1.23	.6	.81	.81	70.0 79.1	30L
15	149	30	.68	.28	1.58	2.0	1.12	.4	.26	.65	60.0 60.7	15P
10	149	30	.68	.28	1.49	1.7	1.16	.5	.49	.65	66.7 60.7	10L
16	151	30	.84	.29	1.29	1.1	1.15	.5	.54	.67	63.3 62.4	16L
5	147	30	.52	.28	1.22	.9	1.17	.6	.45	.64	50.0 55.8	05L
1	144	30	.30	.27	1.09	.4	.88	-.3	.56	.62	60.0 53.6	01L
32	140	30	.02	.26	1.04	.2	1.00	.1	.56	.61	60.0 51.6	32L
20	161	30	1.93	.39	1.02	.2	.96	.2	.85	.83	80.0 79.1	20L
40	149	30	.68	.28	.98	.0	.82	-.4	.62	.65	63.3 60.7	40P
42	164	30	2.46	.47	.98	.0	.63	-.3	.93	.88	80.0 86.1	42L
17	165	30	2.70	.50	.88	.0	.43	-.7	.90	.90	96.7 89.0	17L
4	165	30	2.70	.50	.85	-.1	.37	-.9	.90	.90	96.7 89.0	04L
14	162	30	2.08	.41	.83	-.3	.33	-1.2	.86	.85	90.0 81.8	14L
19	163	30	2.26	.43	.82	-.2	.33	-1.1	.85	.87	86.7 84.6	19P
9	164	30	2.46	.47	.80	-.3	.33	-1.0	.89	.88	90.0 86.1	09L
22	162	30	2.08	.41	.74	-.5	.70	-.3	.92	.85	76.7 81.8	22P
12	164	30	2.46	.47	.74	-.4	.35	-1.0	.89	.88	93.3 86.1	12L
7	165	30	2.70	.50	.72	-.4	.24	-1.2	.91	.90	96.7 89.0	07P
34	164	30	2.46	.47	.66	-.6	.24	-1.3	.90	.88	93.3 86.1	34L
38	165	30	2.70	.50	.66	-.5	.27	-1.1	.93	.90	86.7 89.0	38P
21	162	30	2.08	.41	.61	-.9	.65	-.4	.96	.85	83.3 81.8	21L
45	163	30	2.26	.43	.55	-1.0	.65	-.3	.93	.87	90.0 84.6	45P
23	166	30	2.97	.55	.64	-.5	.33	-.9	.93	.91	90.0 90.3	23L
29	165	30	2.70	.50	.51	-.9	.63	-.3	.96	.90	93.3 89.0	29P
24	164	30	2.46	.47	.60	-.8	.55	-.5	.90	.88	86.7 86.1	24P
35	165	30	2.70	.50	.50	-.9	.57	-.4	.94	.90	93.3 89.0	35P
25	163	30	2.26	.43	.52	-1.1	.52	-.6	.96	.87	90.0 84.6	25P
37	164	30	2.46	.47	.49	-1.1	.51	-.6	.91	.88	93.3 86.1	37P
31	161	30	1.93	.39	.46	-1.5	.37	-1.1	.94	.83	83.3 79.1	31L
11	165	30	2.70	.50	.39	-1.3	.22	-1.3	.94	.90	93.3 89.0	11P
41	165	30	2.70	.50	.39	-1.3	.22	-1.3	.96	.90	93.3 89.0	41L
2	166	30	2.97	.55	.34	-1.3	.35	-.8	.95	.91	96.7 90.3	02L
36	165	30	2.70	.50	.33	-1.5	.18	-1.4	.97	.90	93.3 89.0	36P
3	166	30	2.97	.55	.32	-1.3	.22	-1.2	.95	.91	96.7 90.3	03L
27	166	30	2.97	.55	.31	-1.4	.21	-1.3	.97	.91	96.7 90.3	27P
6	167	30	3.30	.59	.12	-2.0	.05	-2.0	.97	.92	100.0 91.0	06P
8	167	30	3.30	.59	.12	-2.0	.05	-2.0	.97	.92	100.0 91.0	08P
13	167	30	3.30	.59	.12	-2.0	.05	-2.0	.97	.92	100.0 91.0	13L
18	167	30	3.30	.59	.12	-2.0	.05	-2.0	.97	.92	100.0 91.0	18P
33	167	30	3.30	.59	.12	-2.0	.05	-2.0	.97	.92	100.0 91.0	33P
39	167	30	3.30	.59	.12	-2.0	.05	-2.0	.97	.92	100.0 91.0	39P
28	168	30	3.66	.60	.09	-2.0	.05	-2.0	.97	.93	100.0 92.6	28P
MEAN	160.7	30.0	2.20	.45	.74	-.4	.66	-.6			84.3 81.1	
S.D.	8.1	.0	.99	.11	.50	1.3	.84	1.2			17.0 12.7	

Tabel 5b Tabel Guttman Scalograms of Responses – 22.0

GUTTMAN SCALOGRAM OF RESPONSES:

42	+77776767755555555555555555554	42L
19	+77776757555555555555555555555	19P
25	+77777766555555555555555555554	25P
26	+77777776555555555555555555554	26P
45	+77677776655555555555555555554	45P
14	+77777756555555555555555555554	14L
21	+77777776655555555555555555554	21L
22	+77766776555555555555555555554	22P
20	+77777775555555555555555555553	20L
31	+77776676655555555555555555554	31L
30	+77477776555555555555555555554	30L
16	+76657545655555555555555555553	16L
10	+57647475555555555555555555554	10L
15	+54577455555555555555555555553	15P
40	+76746555555555555555555555553	40P
5	+75765455555555555555555555553	05L
1	+75745655555555555555555555553	01L
44	+44665675455555555555555555553	44L
32	+74754555555555555555555555553	32L
43	+77476576654455334533333345444	43L

	1121 111 2222222232 1 11 1	
	433092156625147890078123694785	

Nilai OUTFIT MNSQ, ZSTD dan PT MEASURE CORR. digunakanb untuk kriteria PERSON FIT.

Jumlah $MEAN + S.D. = 0.74 + 0.50 = 1.24$, mengacu kepada Respondance yang memiliki nilai INFIT MNSQ dengan nilai > 1.24 yaitu dari 26P sampai dengan 28P menunjukkan bahwa respondence tersebut adalah FIT.

TABEL 6 TABEL DIMENSIONALITY – 23.0

Tabel 6 Tabel Dimensionality - 23.0

TABLE 23.0 C:\Users\DELL\Desktop\Data pilgan (wi ZOU274WS.TXT) Jun 23 14:44 2026
INPUT: 45 Person 30 Item REPORTED: 45 Person 30 Item 6 CATS WINSTEPS 3.73

Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)				
		-- Empirical --		Modeled
Total raw variance in observations	=	131.5	100.0%	100.0%
Raw variance explained by measures	=	101.5	77.2%	76.4%
Raw variance explained by persons	=	11.1	8.4%	8.3%
Raw Variance explained by items	=	90.4	68.8%	68.1%
Raw unexplained variance (total)	=	30.0	22.8%	23.6%
Unexplned variance in 1st contrast	=	7.2	5.5%	23.9%
Unexplned variance in 2nd contrast	=	3.2	2.4%	10.6%
Unexplned variance in 3rd contrast	=	2.8	2.1%	9.4%
Unexplned variance in 4th contrast	=	2.3	1.7%	7.5%
Unexplned variance in 5th contrast	=	2.1	1.6%	6.9%

STANDARDIZED RESIDUAL VARIANCE SCREE PLOT

Dimensionality atau tidak penelitian ini dapat dilihat dari *Raw Variance Explained by Measure* dan *Unexplained Variance in First Contrast*, mencakup, *Criteria Dimensionality*, yakni nilai *Raw Variance Explained by Measure* sebesar 20%. Hasil menunjukan sebesar 77.2% artinya memenuhi *Dimensionality* (sebaran item memang mengukur construct mindful speech) dengan parameter $> 40\%$ artinya lebih bagus/berhasil dan jika $> 60\%$, artinya Istimewa. *Unexplained Variance*, yakni Varians yang tidak dapat dijelaskan oleh Instrument dan Idealnya tidak melebihi 15%. Berdasarkan Output semuanya $< 15\%$ dan yang terbesar adalah 5.5%. Berdasarkan kriteria tersebut maka dapat disimpulka instrumen yang dikembangkan memenuhi *Dimensiaonality*.

TABEL 7 RATING (PARTIAL CREDIT) SCALE – 3.2

Tabel 7 Rating (Partial Credit) Scale - 3.2

TABLE 3.2 C:\Users\DELL\Desktop\Data pilgan (wit ZOU274WS.TXT, Jun 23 14:44 2026
INPUT: 45 Person 30 Item REPORTED: 45 Person 30 Item 6 CATS WINSTEPS 3.73

SUMMARY OF CATEGORY STRUCTURE. Model="R"

CATEGORY	OBSERVED	OBSVD	SAMPLE	INFIT	OUTFIT	ANDRICH	CATEGORY			
LABEL	SCORE	COUNT	%	AVRGE	EXPECT	MNSQ	MNSQ	THRESHOLD	MEASURE	
2	2	1	0	.95	-2.36	2.37	9.90	NONE	(-6.46)	2
3	3	27	2	-1.76*	-1.70	.98	.56	-5.33	-3.93	3
4	4	76	6	.08	-.57	1.76	.64	-2.21	-2.20	4
5	5	899	67	.94	1.05	.71	.61	-2.27	1.44	5
6	6	79	6	5.32	4.79	.64	.35	5.17	4.90	6
7	7	268	20	6.52	6.49	.92	.71	4.63	(6.30)	7

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

Tabel 7 Rating (Partial Credit) Scale - 3.2 di atas menunjukkan bahwa nilai LOGIT dimulai dari 0.95 (untuk opsi response 2), turun -1.76 > naik 0.08 > naik 0.94 > naik 5.32 > naik 6.52 (untuk opsi response 7) menunjukkan bahwa opsi response (2-7) tidak membingungkan response, artinya response dapat membedakan dengan baik. Selain menggunakan cara *Obsvd Average*, maka dapat juga menggunakan ukuran lainnya, yaitu sebagai berikut, Ukuran lain yakni *Andrich Threshold*, Kriterianya harus bersifat pergerakan dari *None* > *Negative* > *Positif* secara berurutan. Hasil pada tabel tersebut di atas menunjukkan kriteria yang bergerak "Sesuai" pada *None* > *Negative* > *Positif* yang berarti opsi response tersebut adalah *Valid*.

TABEL 8 TABEL DIFFERENTIAL ITEM FUNCTION – 30.0

TABEL 8.A DIF CLASS SPECIFICATION – DIF = \$S3W1

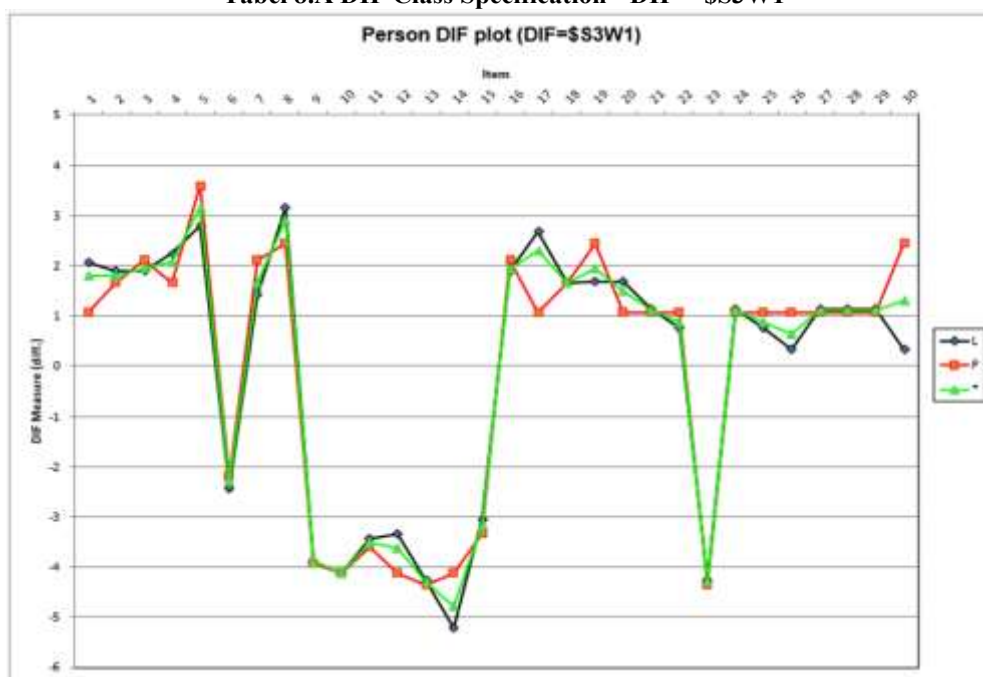
Pada Differential Item Function (DIF), dapat terlihat bahwa apakah Item dalam instrumen bias (menguntungkan kelompok dengan karakteristik tertentu atau tidak). Hal ini dapat terlihat dan teruji dengan ketentuan, jika < 5% dan (< 0,05) maka akan terjadi bias. Terlihat pada Tabel 8a dan pada Chart 8b di bawah ini, berdasarkan pengujian Differential Item Function (DIF) pada kolom Gender, terlihat bahwa Item No. 30 (S30) terlihat memiliki Nilai Probability < 0.05, yaitu: 0.0162 maka dengan demikian Item tersebut termasuk dalam kategori "bias".

Tabel 8 Tabel Differential Item Function - 30.0

DIF class specification is: DIF=\$S3W1

Person CLASSES	SUMMARY DIF CHI-SQUARE	D.F.	PROB.	BETWEEN-CLASS MEAN-SQUARE	t-ZSTD	Item Number Name
2	1.2392	1	.2656	.5182	.0540	1 S1
2	.0779	1	.7802	.0344	-.9599	2 S2
2	.0762	1	.7826	.0321	-.9756	3 S3
2	.4987	1	.4801	.2200	-.3693	4 S4
2	2.3276	1	.1271	1.1688	-.5846	5 S5
2	.3126	1	.5761	.1591	-.5004	6 S6
2	.6656	1	.4146	.2884	-.2485	7 S7
2	1.2220	1	.2690	.5999	.1391	8 S8
2	.0033	1	.9543	.0017	-1.3951	9 S9
2	.0000	1	1.0000	.0006	-1.4727	10 S10
2	.0968	1	.7557	.0483	-.8771	11 S11
2	1.8783	1	.1705	.9683	.4488	12 S12
2	.0196	1	.8886	.0086	-1.2159	13 S13
2	2.7503	1	.0972	1.2860	.6569	14 S14
2	.3846	1	.5351	.1971	-.4154	15 S15
2	.0762	1	.7826	.0321	-.9756	16 S16
2	3.2802	1	.0701	1.4037	.7253	17 S17
2	.0000	1	1.0000	.0006	-1.4605	18 S18
2	1.0856	1	.2974	.4766	.0071	19 S19
2	.4359	1	.5091	.1899	-.4306	20 S20
2	.0051	1	.9428	.0025	-1.3623	21 S21
2	.0856	1	.7698	.0415	-.9155	22 S22
2	.0196	1	.8886	.0086	-1.2159	23 S23
2	.0051	1	.9428	.0025	-1.3623	24 S24
2	.0856	1	.7698	.0415	-.9155	25 S25
2	.4942	1	.4821	.2442	-.3241	26 S26
2	.0051	1	.9428	.0025	-1.3623	27 S27
2	.0051	1	.9428	.0025	-1.3623	28 S28
2	.0051	1	.9428	.0025	-1.3623	29 S29
2	5.7765	1	.0162	2.6222	1.2753	30 S30

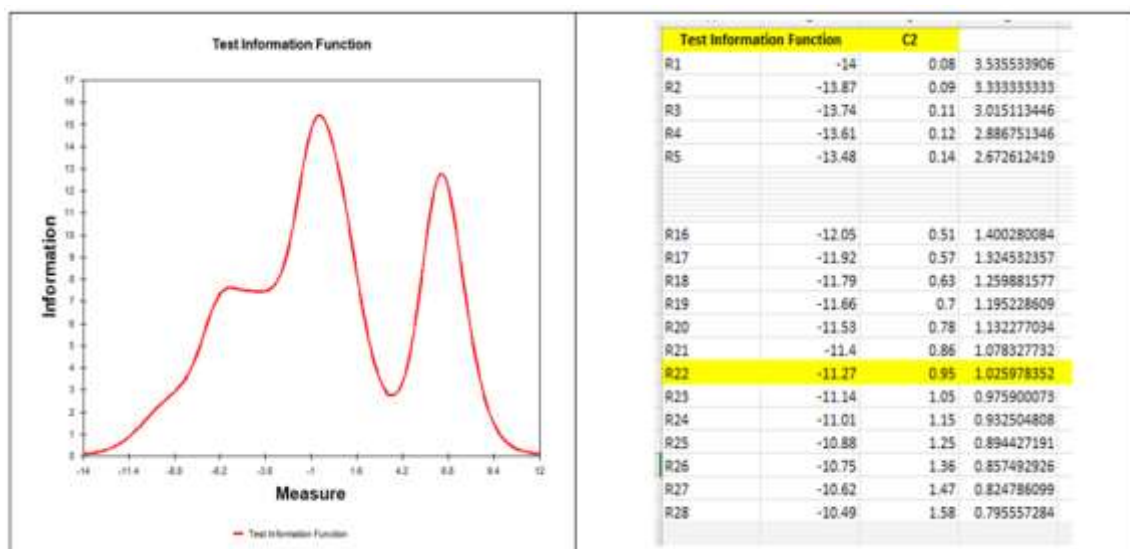
Tabel 8.A DIF Class Specification - DIF = SS3W1



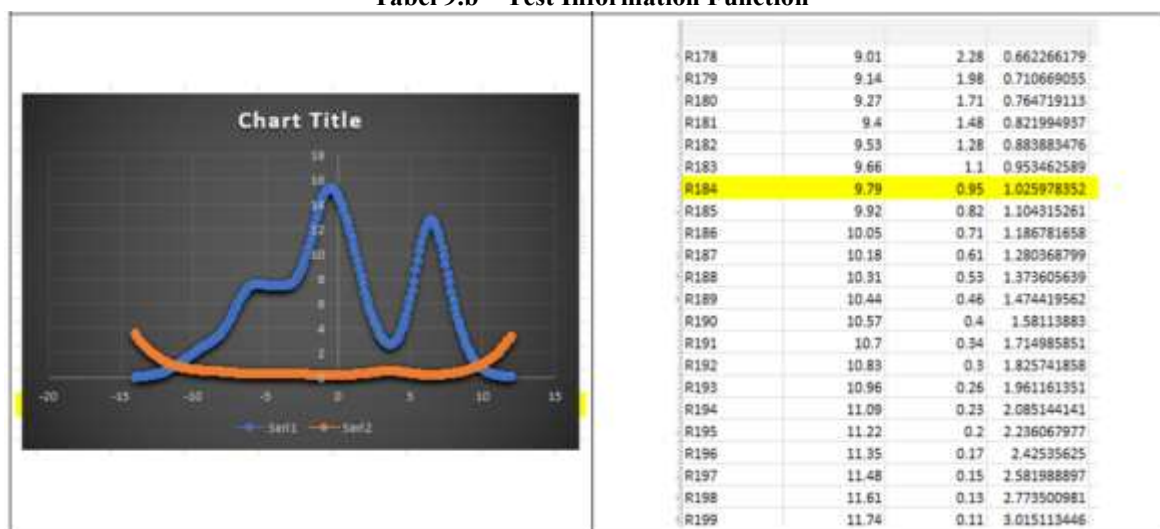
TABEL 9 CHARTS TEST INFORMATION FUNCTION

Dengan menggunakan Charts Test Information Function, dapat menunjukan dengan jelas yang ideal untuk mengukur responden dengan abilitas, sebagai berikut:

Tabel 9.a – Test Information Funtion



Tabel 9.b – Test Information Function

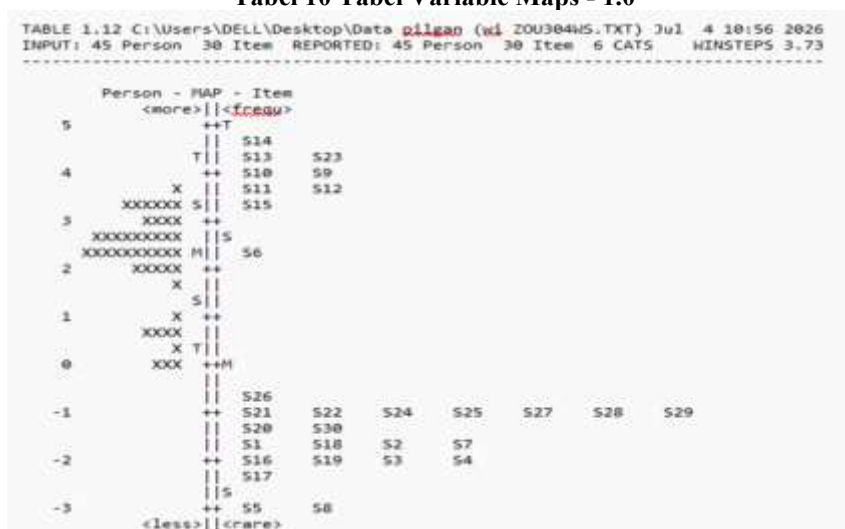


Berdasarkan OUTPUT TEST INFORMATION FUNCTION, instrument memberikan informasi yang ideal jika digunakan untuk mengukur responden dengan abilitas antara -11.27 sampai dengan 9.79.

TABEL 10 TABEL 1. VARIABLE MAPS – 1.0

Analisa *Variable Right Maps* ini membuat dalam suatu skala yang sama kemudian membandingkan sebelah kiri adalah abilitas siswa response dan sebelah kanan adalah Tingkat kesulitan items.

Tabel 10 Tabel Variable Maps - 1.0



Terlihat bahwa Item/Soal paling sulit adalah soal no. 14 sementara item/ soal paling mudah itu adalah item/soal no. 5 dan no. 8. Artinya: Pelajar kelas 11, paling sulit/rendah ketika menjawab item-item pertanyaan no. 14. Bisa dilihat pada soal no. 14 item/soalnya seperti apa, berarti soal itu yang paling sulit dijawab oleh siswa kelas 11. Sementara item/soal no. 5 dan no. 8 ini, paling mudah dijawab oleh pelajar kelas 11. Artinya, pernyataan-pernyataan dalam soal tersebut mudah dipahami/dijawab oleh para pelajar kelas 11. Setelah melakukan penjabaran dan pembahasan atas penelitian yang menggunakan model ADDIE (Analysis, Design, Development, Implementation, Evaluation) dengan pengolahan data SIMULASI ANALISIS DATA POLITOMI (SKALA MINDFULL SPEECH) PEMODELAN RASCH atas Data Analisis Soal Hasil Penilaian Harian Pendidikan Agama Buddha Tahun Pelajaran 2025/2026 pada kelas XI, maka kami pun melakukan Hasil Validasi Ahli (Expert Judgment) dari dua orang Validator, yaitu Ibu Puji Hartati, S. Ag., M. Pd. B.B. dan Validator Bapak Agung Sadha, S. Dt. B., M. Pd., dengan hasil sebagai berikut Hasil Validasi Ahli (Expert Judgment) dari Ibu Puji Hartati, S. Ag., M. Pd.B Proses validasi instrumen penilaian hasil belajar Pendidikan Agama Buddha untuk kelas XI dalam bentuk soal pilihan ganda kompleks (total 30 butir soal) telah dilakukan oleh Pakar/Panelis kedua, yaitu Ibu Puji Hartati, S. Ag., M. Pd. B.B. sebagai Validator melakukan penelaahan komprehensif pada tiga aspek utama: Materi, Konstruksi, dan Bahasa/Budaya menggunakan rentang skala 1 – 4 , yaitu: 1 = (Kurang), 2 = (Cukup), 3 = (Baik) dan 4 = (Sangat Baik). Dan Hasil Validasi Ahli (Expert Judgment) dari Bapak Agung Sadha, S. Dt. B., M. Pd. Proses validasi instrumen penilaian hasil belajar Pendidikan Agama Buddha untuk kelas XI berbentuk pilihan

ganda kompleks (total 30 butir soal) dilakukan oleh seorang pakar, yaitu Bapak Agung Sadha, S. Dt. B., M. Pd.. Pada proses Evaluasi dilakukan dengan menelaah tiga aspek utama, yaitu: Materi, Konstruksi, dan Bahasa/Budaya menggunakan skala penilaian dari angka 1 – 4, yaitu: 1 = (Kurang), 2 = (Cukup), 3 = (Baik) dan 4 = (Sangat Baik).

IV. SIMPULAN

Berdasarkan hasil penelitian dan pembahasan mengenai pengembangan instrumen evaluasi pembelajaran Pendidikan Agama Buddha jenjang SMA (kelas XI) berbasis kontekstual dan HOTS pada Kurikulum Merdeka dengan pendekatan metode pilihan ganda kompleks, dapat disimpulkan sebagai berikut, Kualitas Instrumen Berdasarkan Pemodelan *Rasch*. Penelitian ini berhasil mengembangkan instrumen evaluasi yang memenuhi standar kualitas psikometrik yang baik, sebagaimana ditunjukkan oleh hasil analisis menggunakan pemodelan *Rasch*, Reliabilitas Instrumen. Reliabilitas person sebesar 0,77 (*kategori kuat*), menunjukkan konsistensi jawaban siswa yang baik, Reliabilitas item sebesar 0,98 (*kategori istimewa*), menunjukkan kualitas butir soal yang sangat baik dan Reliabilitas interaksi *person-item* sebesar 0,89 (*kategori baik*), menunjukkan konsistensi internal yang tinggi selanjutnya adalah Validitas Konstruksi, Instrumen memenuhi kriteria dimensionalitas dengan nilai Raw Variance Explained by Measure sebesar 77,2% (*kategori istimewa*, >60%) dan Unexplained variance tertinggi hanya 5,5% (<15%), menunjukkan instrumen mengukur konstruk sikap spiritual secara konsisten. Kesesuaian Model (Model Fit) dengan Nilai INFIT dan OUTFIT MNSQ person (0,74 dan 0,66) mendekati nilai ideal 1,0. Dan Nilai ZSTD (-0,4 dan -0,6) mendekati nilai ideal 0, menunjukkan kesesuaian data dengan model *Rasch*. Validitas Opsi Respons menunjukkan Analisis Rating Scale menunjukkan opsi respons (*skala 2-7 tidak membingungkan responden* dan Nilai OBSVD AVERAGE dan ANDRICH THRESHOLD menunjukkan pola yang valid dan konsisten. Dari 30 butir soal, hanya 1 item (nomor 30) yang teridentifikasi bias berdasarkan *gender* ($p=0,0162$). Instrumen secara keseluruhan dapat digunakan secara adil untuk kedua jenis kelamin.

Karakteristik Instrumen yang Dikembangkan sebagai evaluasi yang dihasilkan, maka memiliki beberapa keunggulan yang dapat disampaikan, sebagai berikut yang pertama adalah Berbasis HOTS dimana dapat Mengukur kemampuan berpikir tingkat tinggi pada level C4 (analisis) dan C5 (evaluasi) dan Mengatasi dominasi soal LOTS dalam praktik evaluasi pembelajaran. Yang kedua adalah Kontekstual dimana Mengintegrasikan konteks budaya lokal dan kehidupan nyata dalam butir soal dan Relevan dengan materi "*Sayangi Seni dan Budaya Agama Buddha*" serta pengembangan inovatif yang Menggunakan model pilihan ganda kompleks yang memungkinkan lebih dari satu jawaban yang benar dan Mampu mengukur kemampuan analisis dan evaluasi secara lebih mendalam dan yang terakhir adalah Terintegrasi dengan Kurikulum Merdeka yang Disusun berdasarkan Capaian Pembelajaran (CP), Tujuan Pembelajaran (TP), dan Alur Tujuan Pembelajaran (ATP).

Setelah melalui pembahasan yang terstruktur, maka berdasarkan hasil analisis *RASCH - Politomy* dan validasi ahli, hipotesis H_1 diterima: Instrumen penilaian sikap spiritual yang dikembangkan terbukti valid dan reliabel. Instrumen ini layak digunakan untuk mengukur sikap spiritual peserta didik kelas XI pada mata pelajaran Pendidikan Agama Buddha dan Budi Pekerti dengan materi "*Sayangi Seni dan Budaya Agama Buddha*" dalam konteks implementasi Kurikulum Merdeka. Instrumen yang dikembangkan menggunakan model *ADDIE* telah melalui tahapan sistematis dan menghasilkan produk evaluasi yang memenuhi standar psikometrik tinggi, mampu mengukur HOTS, kontekstual, dan terintegrasi dengan komponen Kurikulum Merdeka. Dengan demikian, instrumen ini dapat menjadi kontribusi nyata dalam meningkatkan kualitas evaluasi pembelajaran Pendidikan Agama Buddha di Indonesia.

V. REFERENSI

- Arikunto, S. (2013). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Arikunto, S. (2013). *Prosedur penelitian: Suatu pendekatan praktik*. Jakarta: Rineka Cipta.
- Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing*. New York: Longman.
- Cahyadi, R. A. H. (2020). Pengembangan bahan ajar berbasis ADDIE. *Jurnal Pendidikan*.
- Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education*, 5(1), 7–74.
- Branch, R. M. (2020). *Instructional Design: The ADDIE Approach*.
- Buddharakkhita, A. (1985). *The Dhammapada: The Buddha's Path of Wisdom*. Kandy: Buddhist Publication Society.
- Benjamin D. Wright & Geofferey N. Masters (1982). *Rating Scale Analysis (RASCH Measurement)*, Chicago, IL: Mesa Press.
- Emmons, R. A. (2000). Spirituality and intelligence: Problems and prospects. *International Journal for the Psychology of Religion*, 10(1), 57–64.
- Gardner, H. (1999). *Intelligence reframed: Multiple intelligences for the 21st century*. New York: Basic Books.
- Haryanto. (2020). Evaluasi Pembelajaran dalam Meningkatkan Kualitas Pendidikan. *Jurnal Pendidikan Indonesia*, 1(2), 85–94.
- Istiyono, E., Mardapi, D., & Suparno. (2014). Pengembangan Tes Kemampuan Berpikir Tingkat Tinggi Fisika (PYSTHOT). *Jurnal Penelitian dan Evaluasi Pendidikan*, 18(1), 1–12.
- Larry R. Price (2017). *Psychometric Methods (Theory into Practice)*, London: The Guilford Press.
- Lickona, T. (1991). *Educating for character: How our schools can teach respect and responsibility*. New York: Bantam Books.
- Mardapi, D. (2012). *Pengukuran, Penilaian, dan Evaluasi Pendidikan*. Yogyakarta: Nuha Medika.
- Matilde Bini, Paola Monari, Domenico Piccolo & Luigi Salmaso (2009). *Statistical Methods for The Evaluation of Educational Service and Quality of Products*, Italy: Physica-Verlag (A Springer Company).

- Narada Thera. (1998). *The Buddha and His Teachings*. Kuala Lumpur: Buddhist Missionary Society.
- Nyanaponika Thera. (1962). *The heart of Buddhist meditation*. London: Rider & Company.
- Rahula, W. (1974). *What the Buddha taught*. New York: Grove Press.
- Retnawati, H., et al. (2018). Teachers' Knowledge about HOTS. *Problems of Education in the 21st Century*, 76(2), 215–230.
- Setiawan, A., et al. (2020). *Metodologi penelitian pendidikan*. Yogyakarta: Deepublish.
- Setiawan, A., et al. (2021). Metode penelitian pendidikan. *Jurnal Penelitian Pendidikan*.
- Sudjana, N. (2017). *Penilaian Hasil Proses Belajar Mengajar*. Bandung: Remaja Rosdakarya.
- Sugiyono. (2017). Metode penelitian kuantitatif, kualitatif, dan R&D. Bandung: Alfabeta.
- Sugiyono. (2022). *Metode Penelitian Pendidikan*. Bandung: Alfabeta.
- Suryani, N. (2020). Pembelajaran Berbasis Budaya dalam Meningkatkan Karakter Siswa. *Jurnal Pendidikan dan Kebudayaan*, 10(1), 45–56.
- Widana, I. W. (2017). Higher Order Thinking Skills Assessment (HOTS). *Jurnal Pendidikan Indonesia*, 6(2), 150–160.
- Zohar, D., & Marshall, I. (2000). *Spiritual intelligence: The ultimate intelligence*. London: Bloomsbury.